

证 明

9/787238

PCT/CN 99/00139

REC'D 22 SEP 1999

WIPO PCT

本证明之附件是向本局提交的下列专利申请副本

申 请 日: 98 09 22

申 请 号: 98 1 19758.2

申 请 类 别: 发 明

发 明 创 造 名 称: 新的人肝癌细胞衍生生长因子编码序列、
其编码的多肽及制备方法

发明人或设计人: 余龙 张宏来 傅强 赵勇 屠强

申 请 人: 复旦大学

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

中 华 人 民 共 和 国

国家知识产权局局长

姜 颖

99 年 09 月 08 日



1. 一种分离出的DNA分子, 其特征在于, 它包括: 编码具有人HDGF2蛋白活性的多肽的核苷酸序列,

5 所述的核苷酸序列与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列有至少70%的同源性; 或者

所述的核苷酸序列能在中度严紧条件下与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列杂交。

10 2.如权利要求1所述的DNA分子, 其特征在于, 所述的序列编码一多肽, 该多肽具有SEQ ID NO.4所示的序列。

3.如权利要求1所述的DNA分子, 其特征在于, 该序列具有SEQ ID NO.3中核苷酸121-732位的序列。

4.一种分离的HDGF2蛋白多肽, 其特征在于, 它包括: 具有SEQ ID NO.4氨基酸序列的多肽、或其活性片段, 或其活性衍生物。

15 5.如权利要求4所述的多肽, 其特征在于, 该多肽是具有SEQ ID NO.4序列的多肽。

6.一种载体, 其特征在于, 它含有权利要求1所述的DNA。

7.一种用权利要求6所述载体转化的宿主细胞。

8.如权利要求7所述的宿主细胞, 其特征在于, 该细胞是大肠杆菌。

20 9.如权利要求7所述的宿主细胞, 其特征在于, 该细胞是真核细胞。

10.一种产生具有HDGF2蛋白活性的多肽的方法, 其特征在于, 该方法包括:

25 (a)将编码具有HDGF2蛋白活性的多肽的核苷酸序列可操作地连于表达调控序列, 形成HDGF2蛋白表达载体, 所述的核苷酸序列与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列有至少70%的同源性;

(b)将步骤(a)中的表达载体转入宿主细胞, 形成HDGF2蛋白的重组细胞;

(c)在适合表达HDGF2蛋白多肽的条件下, 培养步骤(b)中的重组细胞;

(d)分离出具有HDGF2蛋白活性的多肽。

11.如权利要求10所述的方法, 其特征在于, 该序列为SEQ ID NO.3中从核苷酸121-732位。

30 12.一种能与权利要求4所述的HDGF2蛋白多肽特异性结合的抗体。

13.一种核苷酸分子, 其特征在于, 它是权利要求1所述DNA分子的反义序列。

14. 一种探针分子，其特征在于，它含有权利要求1所述的 DNA 分子中约 8-100 个连续核苷酸。

新的人肝癌细胞衍生生长因子编码序列、
其编码的多肽及制备方法

5

本发明涉及基因工程领域，具体地，本发明涉及一种新的人基因核苷酸序列。更具体地说，本发明涉及新的II型人肝癌细胞衍生生长因子(HDGF2)的cDNA序列，该蛋白是I型HDGF的同系物。本发明还涉及由该核苷酸序列编码的多肽，这些多核苷酸和多肽的应用，以及所述多核苷酸和所述多肽的生产方法。

10

研究已表明，细胞生长的调节是通过各种细胞因子与其特异的膜表面受体作用后引发的一系列级联反应实现的。在肿瘤细胞中，某些级联反应的失控致使细胞持续增殖。在肝癌细胞中，人们已经发现一些自分泌和旁分泌细胞因子(Proc. Natl. Acad. Sci. 83:2448-2452, 1986; Proc. Natl. Acad. Sci. 86:7432-7436, 1989; Cell 61:1137-1146, 1990)。肝癌细胞衍生生长因子(HDGF)是在无血清培养的人肝癌细胞株系HuH-7中找到的一个细胞因子，它具有肝素结合能力并能刺激Swiss 3T3细胞的DNA合成(J. Biol. Chem. 269(40):25143-25149, 1994)。

15

1989年Nakamura等人首次从HuH-7细胞中部分纯化并鉴定了HDGF(Clin. Chim. Acta. 183:273-284, 1989)，1994年该实验组又完整克隆了HDGF的cDNA序列(J. Biol. Chem. 269(40):25143-25149, 1994)，1997年该实验组在小鼠中找到了HDGF的同系物并发现了该基因家族的另两个成员HRP-1、HRP-2，它们都具有一个相当保守的98个氨基酸长的氨基端序列(Biochem. Biophys. Res. Commun. 238:26-32, 1997)。在本发明被公布之前，尚没有任何人公开过本申请中涉及的另一人类HDGF家族成员人HDGF2。

20

25

本发明的一个目的是提供一种新的多核苷酸序列，该多核苷酸序列编码HDGF同源蛋白，本发明的HDGF同源基因被命名为人HDGF2。

本发明的另一个目的是提供一种新的蛋白，该蛋白被命名为人HDGF2。

30

本发明的再一个目的是提供一种利用重组技术生产所述的新的的人HDGF2蛋白的方法。

本发明还提供了这种人HDGF2基因序列和蛋白的应用。

在本发明的一个方面，提供了一种分离出的DNA分子，它包括：编码具有人HDGF2蛋白活性的多肽的核苷酸序列，所述的核苷酸序列与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列有至少70%的同源性；或者所述的核苷酸序列能在中度严紧条件下与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列杂交。较佳地，所述的序列编码一多肽，该多肽具有SEQ ID NO.4所示的序列。更佳地，该序列具有SEQ ID NO.3中从核苷酸121-732位的核苷酸序列。

在本发明的另一方面，提供了一种分离的HDGF2蛋白多肽，它包括：具有SEQ ID NO.4氨基酸序列的多肽、或其活性片段，或其活性衍生物。较佳地，该多肽是具有SEQ ID NO.4序列的多肽。

在本发明的另一方面，提供了一种载体，它含有上述分离出的DNA。

在本发明的另一方面，提供了一种所述载体转化的宿主细胞。

在本发明的另一方面，提供了一种产生具有HDGF2蛋白活性的多肽的方法，该方法包括：

(a)将编码具有HDGF2蛋白活性的多肽的核苷酸序列可操作地连于表达调控序列，形成HDGF2蛋白表达载体，所述的核苷酸序列与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列有至少70%的同源性；

(b)将步骤(a)中的表达载体转入宿主细胞，形成HDGF2蛋白的重组细胞；

(c)在适合表达HDGF2蛋白多肽的条件下，培养步骤(b)中的重组细胞；

(d)分离出具有HDGF2蛋白活性的多肽。

在本发明的一个具体实施方案中，本发明的分离的多核苷酸全长为1024个核苷酸，其详细序列见SEQ ID NO.3，其中开放读框位于121-732位核苷酸。

在本发明中，“分离的”、“纯化的”或“基本纯的”DNA是指，该DNA或片段已从天然状态下位于其两侧的序列中分离出来，还指该DNA或片段已经与天然状态下伴随核酸的组份分开，而且已经与在细胞中伴随其的蛋白质分开。

在本发明中，术语“HDGF2蛋白(或多肽)编码序列”指编码具有HDGF2蛋白活性的多肽的核苷酸序列，如SEQ ID NO.3中121-732位核苷酸序列及其简并序列。该简并序列是指，位于SEQ ID NO.3序列的编码框121-732位核苷酸中，有一个或多个密码子被编码相同氨基酸的简并密码子所取代后而产生的序列。由于密码子的简并性，所以与SEQ ID NO.3中121-732位核苷酸序列同源性低至约70%的简并序列也能编码出SEQ ID NO.4所述的序列。该术语还包括能在中度严紧条件下，更佳地，在高度严紧条件下与SEQ ID NO.3中从核苷酸121-732位的核苷酸序列杂交的核苷酸序列。此外，该术语还包括与SEQ ID NO.3中从核苷酸121-732位

核苷酸序列的同源性至少70%，较佳地至少80%，更佳地至少90%的核苷酸序列。

该术语还包括能编码具有与人HDGF2相同功能的蛋白的、SEQ ID NO.3中开放读框序列的变异形式。这些变异形式包括(但并不限于)：若干个(通常为1-90个，较佳地1-60个，更佳地1-20个，最佳地1-10个)核苷酸的缺失、插入和/或取代，以及在5'和/或3'端添加数个(通常为60个以内，较佳地为30个以内，更佳地为10个以内，最佳地为5个以内)核苷酸。

在本发明中，“基本纯的”蛋白质或多肽是指其至少占样品总物质的至少20%，较佳地至少50%，更佳地至少80%，最佳地至少90%(按干重或湿重计)。纯度可以用任何合适的方法进行测量，如用柱层析、PAGE或HPLC法测量多肽的纯度。基本纯的多肽基本上不含天然状态下的伴随其的组分。

在本发明中，术语“HDGF2蛋白多肽”指具有HDGF2蛋白活性的SEQ ID NO.4序列的多肽。该术语还包括具有与人HDGF2相同功能的、SEQ ID NO.4序列的变异形式。这些变异形式包括(但并不限于)：若干个(通常为1-50个，较佳地1-30个，更佳地1-20个，最佳地1-10个)氨基酸的缺失、插入和/或取代，以及在C末端和/或N末端添加一个或数个(通常为20个以内，较佳地为10个以内，更佳地为5个以内)氨基酸。例如，在本领域中，用性能相近或相似的氨基酸进行取代时，通常不会改变蛋白质的功能。又比如，在C末端和/或N末端添加一个或数个氨基酸通常也不会改变蛋白质的功能。该术语还包括HDGF2蛋白的活性片段和活性衍生物。

该多肽的变异形式包括：同源序列、等位变异体、天然突变体、诱导突变体、在高或低的严谨度条件下能与HDGF2 DNA 杂交的DNA所编码的蛋白、以及利用抗HDGF2多肽的抗血清获得的多肽或蛋白。本发明还提供了其他多肽，如包含HDGF2多肽或其片段的融合蛋白。除了几乎全长的多肽外，本发明还包括了HDGF2多肽的可溶性片段。通常，该片段具有HDGF2多肽序列的至少约10个连续氨基酸，通常至少约30个连续氨基酸，较佳地至少约50个连续氨基酸，更佳地至少约80个连续氨基酸，最佳地至少约100个连续氨基酸。

发明还提供HDGF2蛋白或多肽的类似物。这些类似物与天然HDGF2多肽的差别可以是氨基酸序列上的差异，也可以是不影响序列的修饰形式上的差异，或者兼而有之。这些多肽包括天然或诱导的遗传变异体。诱导变异体可以通过各种技术得到，如通过辐射或暴露于诱变剂而产生随机诱变，还可通过定点诱变法或其他已知分子生物学的技术。类似物还包括具有不同于天然L-氨基酸的残基(如D-氨基酸)的类似物，以及具有非天然存在的或合成的氨基酸(如 β 、 γ -氨基酸)的类

似物。应理解，本发明的多肽并不限于上述例举的代表性的多肽。

修饰(通常不改变一级结构)形式包括：体内或体外的多肽的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化，如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽。这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸，磷酸丝氨酸，磷酸苏氨酸)的序列。还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的多肽。

本发明还包括HDGF2多肽编码序列的反义序列，这种反义序列可用于抑制细胞内HDGF2的表达。

本发明还包括可用作探针和引物的核酸分子，该分子通常具有HDGF2多肽编码序列的8-100个，较佳地15-50个连续核苷酸。该探针可用于检测样品中是否存在编码HDGF2的核酸分子。

本发明还包括检测HDGF2核苷酸序列的方法，它包括用上述的探针与样品进行杂交，然后检测探针是否发生了结合。较佳地，该样品是PCR扩增后的产物，其中PCR扩增引物对应于HDGF2多肽的编码序列，并可位于该编码序列的两侧或中间。引物长度一般为20-50个核苷酸。

在本发明中，可选用本领域已知的各种载体，如市售的载体。

在本发明中，术语“宿主细胞”包括原核细胞和真核细胞。常用的原核宿主细胞的例子包括大肠杆菌、枯草杆菌等。常用的真核宿主细胞包括酵母细胞，昆虫细胞、和哺乳动物细胞。较佳地，该宿主细胞是真核细胞，如CHO细胞、COS细胞等。

另一方面，本发明还包括对HDGF2 DNA或是其片段编码的多肽具有特异性的多克隆抗体和单克隆抗体，尤其是单克隆抗体。这里，“特异性”是指抗体能结合于HDGF2基因产物或片段。较佳地，指那些能与HDGF2基因产物或片段结合但不识别和结合于其它非相关抗原分子的抗体。本发明中抗体包括那些能够结合并抑制HDGF2蛋白的分子，也包括那些并不影响HDGF2蛋白功能的抗体。本发明还包括那些能与修饰或未经修饰形式的HDGF2基因产物结合的抗体。

本发明不仅包括完整的单克隆或多克隆抗体，而且还包括具有免疫活性的抗体片段，如Fab'或(Fab)₂片段；抗体重链；抗体轻链；遗传工程改造的单链Fv分子(Ladner等人，美国专利No. 4,946,778)；或嵌合抗体，如具有鼠抗体结合特异性但仍保留来自人的抗体部分的抗体。

本发明的抗体可以通过本领域内技术人员已知的各种技术进行制备。例如，

纯化的HDGF2基因产物或者其具有抗原性的片段,可被施用于动物以诱导多克隆抗体的产生。与之相似的,表达HDGF2或其具有抗原性的片段的细胞可用来免疫动物来生产抗体。本发明的抗体也可以是单克隆抗体。此类单克隆抗体可以利用杂交瘤技术来制备(见 Kohler 等人, Nature 256:495, 1975; Kohler 等人, Eur.J.Immunol. 6:511, 1976; Kohler 等人, Eur.J.Immunol. 6:292, 1976; Hammerling 等人, In Monoclonal Antibodies and T Cell Hybridomas, Elsevier, N.Y., 1981)。本发明的抗体包括能阻断HDGF2功能的抗体以及不影响HDGF2功能的抗体。本发明的各类抗体可以利用HDGF2基因产物的片段或功能区,通过常规免疫技术获得。这些片段或功能区可以利用重组方法制备或利用多肽合成仪合成。与HDGF2基因产物的未修饰形式结合的抗体可以用原核细胞(例如*E. Coli*)中生产的基因产物来免疫动物而产生;与翻译后修饰形式结合的抗体(如糖基化或磷酸化的蛋白或多肽),可以用真核细胞(例如酵母或昆虫细胞)中产生的基因产物来免疫动物而获得。

在附图中,

图1为本发明HDGF2与小鼠HDGF2的核酸序列的同源比较图。其中,相同的核苷酸用“|”标出。

图2为本发明HDGF2与小鼠HDGF2的氨基酸序列的同源比较图。其中,相同的氨基酸用“|”标出,相似的氨基酸用“。”标出。

在本发明中,人HDGF2的 cDNA核苷酸序列是如此获得的,以人睾丸λgt11cDNA文库(购自Clontech公司)为模板,合成正向引物A1: 5'-ACCGCTCGTCCGCCCGGCTTGAG-3' 和反向引物A2: 5'-GATCCTAGACATGTATAAGTCTGCG C-3',进行PCR,分别获得1024bp的目的片段。测序后得到SEQ ID NO.3的全长cDNA序列。

肝癌细胞衍生生长因子(HDGF)是从人的肝癌细胞株HuH-7中分离到的肝素结合蛋白,它具有刺激细胞生长的活性,能够促进成纤维细胞和一些肝癌细胞的生长(J. Biol. Chem. 269(40):25143-25149, 1994)。它在人的心、脑、肺、肝等各个组织及各种癌细胞株中均有表达 (J. Biol. Chem. 269(40):25143-25149, 1994)。HDGF家族的已知各成员的表达模式是不同的,但是它们在精巢中都有很高程度的富集,而且它们的5'非翻译区均有高于70%的GC比(Biochem. Biophys. Res. Commun. 238:26-32, 1997),因而可能在雄性生殖细胞发育过程中有重要功能,

和DNA甲基化、染色质构象以及翻译调控相关(J. Cell. Biol. 115:887-903, 1990; Cell 62:503-514, 1990)。尽管HDGF蛋白主要存在于细胞质(J. Biol. Chem. 269(40):25143-25149, 1994), 但是该家族成员的氨基酸序列中都含有一个潜在的核定位信号(NLS), 并且都无信号肽顺序, 提示它们可能作为核蛋白起作用。另外, HDGF的C端的酸性氨基酸尾和HMG家族的HMG-1/-2高度同源, 而这段序列在HMG-1/-2中已知是组蛋白结合区域(Biochemistry 29:4419-4423, 1990)。很有可能HDGF在内化后作为转录因子发挥其刺激细胞生长的活性(Biochem. Biophys. Res. Commun. 238:26-32, 1997)。HDGF的有丝分裂原活性使其在急性恶性肝炎和肝损伤的治疗上存在着极大的应用价值(Clin. Chim. Acta. 183:273-284, 1989)。研究表明, 许多成纤维细胞生长因子能够广泛应用于局部缺血症及动脉粥样硬化症等血管形成方面有缺陷的疾病以及神经细胞的发育(Blood 91(10):3527-3561, 1998; Ann. N. Y. Acad. Sci. 545:240-252, 1998)。

下面结合具体实施例, 进一步阐述本发明。应理解, 这些实施例仅用于说明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法, 通常按照常规条件如Sambrook等人, 分子克隆: 实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件, 或按照制造厂商所建议的条件。

实施例1

HDGF2的cDNA序列的克隆和测定

1. 引物扩增

以人睾丸 λ gt11cDNA文库(购自Clontech公司)为模板, 用寡核苷酸A1: 5'-ACCGCTCGTCCGCCCGGCTTGAG-3' (SEQ ID NO.1)为正向引物, 寡核苷酸A2: 5'-GATCCTAGACATGTATAAGTCTGCGC-3' (SEQ ID NO.2)为反向引物, 进行PCR, PCR条件为93℃ 4分钟, 随之以93℃ 1分钟、68.5℃ 1分钟和72℃ 1分钟进行35个循环, 最后72℃延伸5分钟。电泳检测得到的PCR片段, 为1024bp的目的片段。

2. PCR产物的测序

将如上获得的PCR扩增产物与pGEM-TTM载体(Promega)连接, 转化大肠杆菌JM103, 用QIAprep Plasmid试剂盒(QIAGEN)提取质粒, 用双链嵌套式缺失试剂盒(Pharmacia)对插入片段进行定向系列缺失, 然后用PCR对缺失子进行快速鉴定及

排序。用SequiTherm EXCEL™ DNA测序试剂盒(Epicentre Technologies)对依次截短的缺失子进行测序,最后用电脑软件拼接顺序,获得全长cDNA序列,共1024bp,详细序列见SEQ ID NO.3,其中开放读框位于121-732位核苷酸。

根据得到的全长cDNA序列推导出HDGF2的氨基酸序列,共203个氨基酸残基,其氨基酸序列详见SEQ ID NO.4。

实施例2

同源比较

用人HDGF2的全长cDNA序列及其编码蛋白在Non-redundant GenBank+EMBL +DDBJ+PDB 数据库及 Non-redundant GenBank CDS translations+PDB +SwissProt+Spupdate+PIR数据库中用BLAST进行核酸和蛋白同源检索。结果发现它们与小鼠HDGF(dbj|D63707|MUSHDGF)基因及其编码蛋白具有极高同源性,用PCGENE软件比较发现,它们在核酸水平上的同一性达到了68.7%,在蛋白水平上的同一性达到了53.7%,并还有9.4%的氨基酸相似(图1和图2)。特别是在保守的由98个氨基酸所构成的氨基末端,与小鼠HDGF的同源性高达90%。此外,人HDGF2和另一个小鼠HDGF基因(dbj|D63850|D63850)以及另一个人的HDGF基因(dbj|D16431|HUMHDGF)也有一定的同源性。上述这些基因被认为构成一个家族,所以可以从已知的这些基因或蛋白的功能来推测人HDGF2的功能。

肝癌细胞衍生长因子(HDGF)是从人的肝癌细胞株HuH-7中分离到的肝素结合蛋白,它具有刺激细胞生长的活性,能够促进成纤维细胞和一些肝癌细胞的生长(J. Biol. Chem. 269(40):25143-25149, 1994)。尽管HDGF最初在肝癌细胞中被发现,但Northern杂交分析显示它在人的心、脑、肺、肝等各个组织及各种癌细胞株中均有表达。它在正常细胞和肿瘤细胞中是否存在着表达差异,还需进一步实验证明(J. Biol. Chem. 269(40):25143-25149, 1994)。随着研究的深入,HDGF在肝癌细胞中的作用及其对肝癌治疗的影响也会不断被揭示。HDGF家族已知各成员的表达模式是不同的,但是它们在精巢中都有很高程度的富集,而且这些基因的5'非翻译区均有高于70%的GC比(Biochem. Biophys. Res. Commun. 238:26-32, 1997),这个特点和一些在精巢或胚胎发育中特异表达的基因相类似,因而它们可能在雄性生殖细胞发育过程中有重要功能,并和DNA甲基化,染色质构象以及翻译调控相关(J. Cell. Biol. 115:887-903, 1990; Cell 62:503-514, 1990)。

荧光免疫实验显示HDGF蛋白主要存在于细胞质(J. Biol. Chem. 269(40):25143-25149, 1994),但是该家族成员的氨基酸序列中都含有一段碱性区

一个潜在的位信号(NLS), 并且都无信号肽序, 提示它们可能作为核蛋白起作用。成纤维细胞生长因子(FGF)必须依靠这段信号区域, 使之定位于核内而发挥其有丝分裂原的活性。另外, HDGF的C端的酸性氨基酸尾和HMG家族的HMG-1/-2高度同源, 而这段序列在HMG-1/-2中已知是组蛋白结合区域(Biochemistry 29:4419-4423, 1990)。综合上述现象, 很有可能HDGF在内化后作为转录因子发挥其刺激细胞生长的活性(Biochem. Biophys. Res. Commun. 238:26-32, 1997)。本发明的HDGF2也同样具有类似活性。

HDGF的有丝分裂原活性使其在急性恶性肝炎和肝损伤的治疗上存在着极大的应用价值(Clin. Chim. Acta. 183:273-284, 1989)。研究表明, 许多成纤维细胞生长因子具有促进上皮细胞生长的能力, 能够广泛应用于局部缺血症及动脉粥样硬化症等血管形成方面有缺陷的疾病以及神经细胞的发育(Blood 91(10):3527-3561, 1998; Ann. N. Y. Acad. Sci. 545:240-252, 1998)。I型HDGF及本发明HDGF2的在促成纤维细胞生长活性方面的应用尚有待于进一步研究。

本发明的HDGF2除了可作为该家族一员用于进一步的功能研究, 还可用于与其他蛋白一起产生融合蛋白, 比如与免疫球蛋白一起产生融合蛋白。此外, 本发明HDGF2还可以与该家族的其他成员进行融合或交换片段, 以产生新的蛋白, 如将本发明HDGF2的N端与I型HDGF或鼠HDGF的N端进行交换, 以产生新的活性更高或具有新特性的蛋白。

针对本发明HDGF2的抗体, 用于筛选该家族的其他成员, 或者用于亲和纯化相关蛋白(如该家族的其他成员)。

实施例3

HDGF2在大肠杆菌中的表达

在该实施例中, 将编码HDGF2的cDNA序列用对应于该DNA序列的5'和3'端的PCR寡核苷酸引物进行扩增, 获得HDGF2 cDNA作为插入片段。

PCR反应中使用的5'寡核苷酸引物序列为:

5'-CCACGGATCCATGGCGCGTCCGCGGCCCC-3'(SEQ ID NO.5)

该引物含有BamHI限制性内切酶的酶切位点, 在该酶切位点之后是由起始密码子开始的HDGF2编码序列的19个核苷酸;

3'端引物序列为:

5'-ATCCGTCGACTTAGGTCCCTTCACTGGTT-3'(SEQ ID NO. 6)

该引物含有SalI限制性内切酶的酶切位点、翻译终止子和HDGF2的部分编码

引物上的限制性内切酶的酶切位点对应于细菌表达载体pQE-9(Qiagen Inc., Chatsworth, CA)上的限制性内切酶酶切位点, 该质粒载体编码抗生素抗性(Amp^r)、一个细菌复制起点(ori)、一个IPTG-可调启动子/操纵子(P/O)、一个核糖体结合位点(RBS)、一个6-组氨酸标记物(6-His)以及限制性内切酶克隆位点。

用BamHI和SalI消化pQE-9载体及插入片段, 随后将插入片段连接到pQE-9载体并保持开放读框在细菌RBS起始。随后用连接混合物转化购自Qiagen, 商品名为M15/rep4的*E. coli*菌株, M15/rep4含有多拷贝的质粒pREP4, 其表达lacI阻遏物并携带卡那霉素抗性(Kan^r)。在含有Amp和Kan的LB培养皿上筛选转化子, 抽提质粒, 用PstI酶切鉴定插入片段大小及方向, 并测序验证HDGF2的cDNA片段已正确插入了载体。

在补加Amp(100 μ g/ml)和Kan(25 μ g/ml)的LB液体培养基中过夜培养(O/N)含所需构建物的阳性转化子克隆。过夜(O/N)培养物以1: 100-1: 250的稀释率稀释, 然后接种到大体积培养基中, 培养细胞生长至600光密度(OD₆₀₀)为0.4-0.6时, 加入IPTG(“异丙基硫代- β -D-半乳糖苷”)至终浓度为1mM。通过使lacI阻遏物失活, IPTG诱导启动P/O导致基因表达水平提高。继续培养细胞3-4小时, 随后离心(6000 \times g, 20分钟)。超声裂解包涵体, 收集细胞并将细胞沉淀溶于6M的盐酸胍中。澄清后, 通过在能使含6-His标记物蛋白紧密结合的条件下, 用镍-螯合柱层析从溶液中纯化溶解的HDGF2。用6M盐酸胍(pH5.0)从柱中洗脱HDGF2。可用几种方法从盐酸胍中变性沉淀蛋白。或者使用透析步骤除去盐酸胍, 或者从镍-螯合柱中分离出纯化蛋白, 纯化后的蛋白可以结合到第二个柱中, 该柱中具有递减的线性盐酸胍梯度。在结合到该柱时蛋白质变性, 随后用盐酸胍(pH5.0)洗脱。最后, 将可溶的蛋白质对含PBS进行透析, 然后将蛋白质保存在终浓度为10%(w/v)甘油的贮存液中。

用12%的SDS-PAGE胶进行电泳, 鉴定表达蛋白的分子量大小为约23KDa。

此外, 用常规方法对表达蛋白的N端和C端各10个氨基酸长度的氨基酸进行测序, 发现与SEQ ID NO.4的序列一致。

实施例4

HDGF2在真核细胞(CHO细胞株)中的表达

在该实施例中, 将编码HDGF2的cDNA序列用对应于该DNA序列的5'和3'端的PCR寡核苷酸引物进行扩增, 获得HDGF2 cDNA作为插入片段。

PCR反应中使用5'寡核苷酸引物序列为:

5'-CCCTAAGCTTATGGCGCGTCCGCGGCCCC-3'(SEQ ID NO.7)

该引物含有HindIII限制性内切酶的酶切位点, 在该酶切位点之后是由起始密码子开始的HDGF2编码序列的19个核苷酸;

3'端引物序列为:

5'-TTTCGGATCCTTAGGTCCCTTCACTGGTT-3'(SEQ ID NO.8)

该引物含有BamHI限制性内切酶的酶切位点、一个翻译终止子和HDGF2的部分编码序列。

引物上的限制性内切酶的酶切位点对应于CHO细胞表达载体pcDNA3上的限制性内切酶酶切位点, 该质粒载体编码抗生素抗性(Amp^r和Neo^r)、一个噬菌体复制起点(f1 ori)、一个病毒复制起点(SV40 ori)、一个T7启动子、一个病毒启动子(P-CMV)、一个Sp6启动子、一个SV40启动子、一个SV40加尾信号和相应的polyA顺序、一个BGH加尾信号和相应的polyA顺序。

用HindIII和BamHI消化pcDNA3载体及插入片段, 随后将插入片段连接到pcDNA3载体。随后用连接混合物转化*E.coli* DH5 α 菌株。在含有Amp的LB培养皿上筛选转化子, 在补加Amp(100 μ g/ml)的LB液体培养基中过夜培养(O/N)含所需构建物的克隆。抽提质粒, 用PstI酶切鉴定插入片段大小及方向, 并测序验证HDGF2的cDNA片段已正确插入了载体。

质粒转染CHO细胞是用脂转染法, 用Lipofectin试剂盒(GiBco Life)进行。转染48小时后, 经2-3周的持续G418加压筛选, 收集细胞及细胞上清测定表达蛋白酶活力。去G418, 连续传代培养; 对混合克隆细胞极限稀释, 选择具有较高蛋白酶活性的细胞亚克隆。按常规方法大量培养上述阳性亚克隆。48小时后, 开始收集细胞及上清, 用超声裂解方法破碎细胞。以含0.05%Triton的50mMTris·HCl(pH7.6)溶液为平衡液及洗脱液, 用经预平衡的Superdex G-75柱收集上述蛋白的活性峰。再用50mMTris·HCl(pH8.0)平衡的DEAE-Sepharose柱, 以含0-1M NaCl的50mMTris·HCl(pH8.0)溶液为洗脱液进行梯度洗脱, 收集上述蛋白的活性峰。最后以PBS(pH7.4)为透析液对表达蛋白溶液进行透析。最后冻干保存。

用12%的SDS-PAGE胶进行电泳, 鉴定表达蛋白的分子量大小为23KDa。

此外, 用常规方法对表达蛋白的N端和C端各10个氨基酸长度的氨基酸进行测序, 发现与SEQ ID NO.4的序列一致。

实施例5

制备抗体

将实施例3和4获得的重组蛋白用来免疫动物以产生抗体，具体如下。重组分子用层析法进行分离后备用。也可用SDS-PAGE凝胶电泳法进行分离，将电泳条带从凝胶中切下，并用等体积的完全Freund's佐剂乳化。用50-100 μ g/0.2ml乳化的蛋白，对小鼠进行腹膜内注射。14天后，用非完全Freund's佐剂乳化的同样抗原，对小鼠以50-100 μ g/0.2ml的剂量进行腹膜内注射以加强免疫。每隔14天进行一次加强免疫，至少进行三次。获得的抗血清的特异反应活性用它在体外沉淀HDGF2基因翻译产物的能力加以评估。

在本发明提及的所有文献都在本申请中引用作为参考，就如同每一篇文献被单独引用作为参考那样。此外应理解，在阅读了本发明的上述讲授内容之后，本领域技术人员可以对本发明作各种改动或修改，这些等价形式同样落于本申请所附权利要求书所限定的范围。

(1)一般信息:

5

(i)申请人: 复旦大学

(ii)发明名称: 新的人肝癌细胞衍生生长因子编码序列、
其编码的多肽及制备方法

10

(iii)序列数目: 8

(2)SEQ ID NO.1的信息

(i)序列特征

(A)长度: 23碱基

15

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO.1

20

ACCGCTCGTC CGCCCGGCTT GAG

23

(2)SEQ ID NO.2的信息

(i)序列特征

(A)长度: 26碱基

25

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

(xi)序列描述: SEQ ID NO.2

30

GATCCTAGAC ATGTATAAGT CTGCGC

26

(2)SEQ ID NO.3的信息:

(i)序列特征

(A)长度: 1024bp

(B)类型: 核酸

(C)链性: 单链

(D)拓扑结构: 线性

5 (ii)分子类型: cDNA

(xi)序列描述: SEQ ID NO.3

```
1  ACCGCTCGTC CGCCCGGCTT GAGGCCCGCG GGGAGCGCGC GCAATTCGTC GGCCCGCGGG
61  GGGGCGGCCT CCCGGCATCT TCGCGGCGAC CAAGGACTAC CAGGAAGGGG AGCGGCTGGG
10 121 ATGGCGCGTC CGCGGCCCCG CGAGTACAAA GCGGGCGACC TGGTCTTCGC CAAGATGAAG
181 GGCTACCCGC ACTGGCCGGC CCGGATTGAT GAACTCCCAG AGGGCGCTGT GAAGCCTCCA
241 GCAAACAAGT ATCCTATCTT CTTTTTTGGC ACCCATGAAA CTGCATTTCT AGGTCCCAAA
301 GACCTTTTTTC CATATAAGGA GTACAAAGAC AAGTTTGGAA AGTCAAACAA ACGGAAAGGA
361 TTTAACGAAG GATTGTGGGA AATAGAAAAT AACCCAGGAG TAAAGTTTAC TGGCTACCAG
15 421 GCAATTCAGC AACAGAGCTC TTCAGAAACT GAGGGAGAAG GTGGAAATAC TGCAGATGCA
481 AGCAGTGAGG AAGAAGGTGA TAGAGTAGAA GAAGATGGAA AAGGCAAAAG AAAGAATGAA
541 AAAGCAGGCT CAAAACGGAA AAAGTCATAT ACTTCAAAGA AATCCTCTAA ACAGTCCCGG
601 AAATCTCCAG GAGATGAAGA TGACAAAGAC TGCAAAGAAG AGGAAAACAA AAGCAGCTCT
661 GAGGGTGGAG ATGCGGGCAA CGACACAAGA AACACAACCT CAGACTTGCA GAAAACCACT
20 721 GAAGGGACCT AACTACCATA ATGAATGCTG CATATTAAGA GAAACCACAA GAAGGTTATA
781 TGTTTGGTTG TCTAATATTC TTGGATTGTA TATGAACCAA CACATAGTCC TTGTTGTCAT
841 TGACAGAACC CCAGTTTGTA TGTACATTAT TCATATTCTT CTCTGTTGTG TTTCGGGGGG
901 AAAAGACATT TTAGCCTTTT TAAAAGTTA CTGATTTAAT TTCATGTTAT TTGGTTGCAT
961 GAAGTTGCCC TTAACCACTA AGGATTATCA AGATTTTTCG GCAGACTTAT ACATGTCTAG
25 1021 GATC
```

(2)SEQ ID NO.4的信息.

(i)序列特征

(A)长度: 203个氨基酸

30 (B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型: 多肽

(xi)序列描述 SEQ ID NO.4

1 Met Ala Arg Pro Arg Pro Arg Glu Tyr Lys Ala Gly Asp Leu Val
16 Phe Ala Lys Met Lys Gly Tyr Pro His Trp Pro Ala Arg Ile Asp
5 31 Glu Leu Pro Glu Gly Ala Val Lys Pro Pro Ala Asn Lys Tyr Pro
46 Ile Phe Phe Phe Gly Thr His Glu Thr Ala Phe Leu Gly Pro Lys
61 Asp Leu Phe Pro Tyr Lys Glu Tyr Lys Asp Lys Phe Gly Lys Ser
76 Asn Lys Arg Lys Gly Phe Asn Glu Gly Leu Trp Glu Ile Glu Asn
91 Asn Pro Gly Val Lys Phe Thr Gly Tyr Gln Ala Ile Gln Gln Gln
10 106 Ser Ser Ser Glu Thr Glu Gly Glu Gly Gly Asn Thr Ala Asp Ala
121 Ser Ser Glu Glu Glu Gly Asp Arg Val Glu Glu Asp Gly Lys Gly
136 Lys Arg Lys Asn Glu Lys Ala Gly Ser Lys Arg Lys Lys Ser Tyr
151 Thr Ser Lys Lys Ser Ser Lys Gln Ser Arg Lys Ser Pro Gly Asp
166 Glu Asp Asp Lys Asp Cys Lys Glu Glu Glu Asn Lys Ser Ser Ser
15 181 Glu Gly Gly Asp Ala Gly Asn Asp Thr Arg Asn Thr Thr Ser Asp
196 Leu Gln Lys Thr Ser Glu Gly Thr

(2)SEQ ID NO.5的信息

(i)序列特征

20 (A)长度: 29碱基
(B)类型: 核酸
(C)链性: 单链
(D)拓扑结构: 线性

(ii)分子类型: 寡核苷酸

25 (xi)序列描述: SEQ ID NO.5

CCACGGATCC ATGGCGCGTC CGCGGCCCC

29

(2)SEQ ID NO.6的信息

(i)序列特征

30 (A)长度: 29碱基
(B)类型: 核酸
(C)链性: 单链



- (D)拓扑结构: 线性
- (ii)分子类型: 寡核苷酸
- (xi)序列描述: SEQ ID NO.6

ATCCGTCGAC TTAGGTCCCT TCACTGGTT 29

5

(2)SEQ ID NO.7的信息

- (i)序列特征
 - (A)长度: 29碱基
 - (B)类型: 核酸
 - (C)链性: 单链
 - (D)拓扑结构: 线性

10



- (ii)分子类型: 寡核苷酸
- (xi)序列描述: SEQ ID NO.7

CCCTAAGCTT ATGGCGCGTC CGCGGCCCC 29

15

(2)SEQ ID NO.8的信息

- (i)序列特征
 - (A)长度: 29碱基
 - (B)类型: 核酸
 - (C)链性: 单链
 - (D)拓扑结构: 线性

20



- (ii)分子类型: 寡核苷酸
- (xi)序列描述: SEQ ID NO.8

TTTCGGATCC TTAGGTCCCT TCACTGGTT 29

25

说明书附图

人 HDGF2 核酸	- ACCGCTCGTCCGCCCGGCTTGAGGCCCGCGGGAGCGCGCAATTCGTC -50
小鼠 HDGF 核酸	- CGCAAAC-TTG -10
人 HDGF2 核酸	- GGCCCCGCGGGGGGCGGCCTCCCGGCATCTTCGCGGCGACCAAGGACTAC -100
小鼠 HDGF 核酸	- GGCTCGCGC-----TTCCCGGCT-CGGCGCGGAGCCCGG-GGCGCC -49
人 HDGF2 核酸	- CAGGAAGGGGAGCGGCTGGGATGGCGCG-TCCG--CGGCCCCGCGAGTAC -147
小鼠 HDGF 核酸	- CG-----CGGCCCCGCCA---TGTCGCGATCCAACCGGCAGAAAGAGTAC -91
人 HDGF2 核酸	- AAAGCGGGCGACCTGGTCTTCGCCAAGATGAAGGGCTACCCGCACTGGCC -197
小鼠 HDGF 核酸	- AAGTGCGGAGACCTGGTGTTCGGAAGATGAAAGGATACCCACACTGGCC -141
人 HDGF2 核酸	- GGCCCCGATTGATGAACTCCCAGAGGGCGCTGTGAAGCCTCCAGCAAACA -247
小鼠 HDGF 核酸	- GGCCCCGATTGATGAGATGCCTGAGGCTGCAGTGAAGTCAACAGCCAACA -191
人 HDGF2 核酸	- AGTATCCTATCTTCTTTTTTGGCACCCATGAACTGCATTTCTAGGTCCC -297
小鼠 HDGF 核酸	- AATACCAAGTCTTTTTTTTTGGGACCCATGAGACGGCATTCTGGGCCCC -241
人 HDGF2 核酸	- AAAGACCTTTTTCCATATAAGGAGTACAAAGACAAGTTTGAAAGTCAAA -347
小鼠 HDGF 核酸	- AAAGACCTCTCCCTTATGAGGAATCCAAGGAGAAGTTTGGAAGCCCAA -291
人 HDGF2 核酸	- CAAACGGAAGGATTTAACGAAGGATTGTGGGAAATAGAAAATAACCCAG -397
小鼠 HDGF 核酸	- CAAGAGGAAAGGGTTCAGCGAGGGGCTGTGGGAGATCGAGAACAACCCTA -341
人 HDGF2 核酸	- GAGTAAAGTTTACTGGCTACCAGGCAATTCAGCAACAGAGCTCTTC---A -444
小鼠 HDGF 核酸	- CAGTCAAGGCCTCTGGCTACCAGTCTCCAGAAAAAGAGTTGTGCGGCA -391
人 HDGF2 核酸	- GAAAC-----TGAGGGAGAAGGTGGAAATAC--- -470
小鼠 HDGF 核酸	- GAGCCCGAGGTGGAGCCCGAAGCCCATGAGGGTGACGGTGATAAGAAGGG -441
人 HDGF2 核酸	- ---TGCAGATGCAAGCAGTGAGGAAGAAGG-----TGATAGAGTA-- -507
小鼠 HDGF 核酸	- CAGTGCAGAGGGCAGCAGCGACGAAGAAGGGAAACTGGTGATCGATGAAC -491

图 1

人 HDGF2 核酸	- -----GAAGAAGATGGAAAAGGCCA-----AAGAA-AGA-----AT -537
小鼠 HDGF 核酸	- CAGCCAAGGAGAAGAACGAAAAGGGCACGCTGAAGAGGAGAGCAGGGGAT -541
人 HDGF2 核酸	- G-----AAA-----AAGCAGGCTCAAAAC-----GGA -559
小鼠 HDGF 核酸	- GTGTTGGAGGACTCCCCTAAACGTCCCAAGGAGTCAGGAGACCATGAGGA -591
人 HDGF2 核酸	- AAAAGTCAT----ATA-----CTT-----CA----- -576
小鼠 HDGF 核酸	- GGAGGACAAGGAGATAGCTGCCTTGGAGGGTGAGAGGCACCTGCCTGTAG -641
人 HDGF2 核酸	- -----AAGAA-ATC-----CTCTAAAC-AGTC-----CCGGAATCT -606
小鼠 HDGF 核酸	- AGGTGGAGAAGAACAGCACCCCCTCTGAGCCAGACTCTGGCCAGGGACCT -691
人 HDGF2 核酸	- CCAGGAGATGAAGATGACAAAGA-----CTGCAAAG-AAGAGG---A -644
小鼠 HDGF 核酸	- CCTGCAGAGGAAGAAGAGGGAGAGGAAGAGGCTGCCAAGGAAGAGGCTGA -741
人 HDGF2 核酸	- A-----AA-----CAAA -651
小鼠 HDGF 核酸	- AGCCCCAGGCGTCAGAGATCATGAGAGCCTGTAGCCACCAATGTTTCAAG -791
人 HDGF2 核酸	- AGCAGC-----TCTGAGGG---TGGAGATGCG -675
小鼠 HDGF 核酸	- AGGAGCCCCTGCCCCGTTCTGCTGCTGTCTGGGTGCTACTGGGGAAACT -841
人 HDGF2 核酸	- GGCAACGACA--CAA-----GAA-----ACACAACCT- -699
小鼠 HDGF 核酸	- GGCCATGGCCTGCAAACCTGGGAACCCTTTCCACCCTATTTACCCTACTC -891
人 HDGF2 核酸	- --TCAG--ACT---TGAGAAAACC-AGT---GAAG-----GGACCT -730
小鼠 HDGF 核酸	- CCTCACTCACTCTCTCCTCTAAGCCCACTCCTGGAGAGTGTCTTGGCCCT -941
人 HDGF2 核酸	- AACTACCA-----TA-ATGAATGCTG---CATATTAAGAGA--AA -764
小鼠 HDGF 核酸	- CACCTCCAGCTCCCTTCTATATACACCCTGTGCCCCAGGATGAGATGAG -991
人 HDGF2 核酸	- CCACAAGAAGGT-TATA---TGTTT-----GGTT-----GTCTAA -795
小鼠 HDGF 核酸	- GCCTTTGTATCTCTTTACACTTGTTTCCCAGGGTTTCTGCTGGGGTCTAG -1041
人 HDGF2 核酸	- TAT-----TCTTG-----GA -805
小鼠 HDGF 核酸	- GCTGCTGTTTCCACCTCTTGACACCTCTGCCCTGCTGCAGGCATTCTAGA -1091

图 1(续)

人 HDGF2 核酸	- --TTTG-----ATA-----TGAACCAACACATAG--- -827
小鼠 HDGF 核酸	- CCTTTGGGGTGGATAGTGGGCAGGAGTGGAGGTGAAAGAATATAAAGGAG -1141
人 HDGF2 核酸	- -----TCCTTGTGTGTCATTGAC-----AGAACC-----CCAG----- -854
小鼠 HDGF 核酸	- TGTTGGSTTCATGGATGGCATCGTCTACCTGAGCTCCTGTCTCCAGCCCCC -1191
人 HDGF2 核酸	- -----TTTG----TATG--TACATT----- -868
小鼠 HDGF 核酸	- ACACTTATTTTCCCATCTGCCTACATTCAAGAAACAGGACACTGTGGGAG -1241
人 HDGF2 核酸	- -----ATTCAT--ATTCCTCTCTGTTGTGTTTCGGG----- -897
小鼠 HDGF 核酸	- AGAGGCTACCATCCATCCATAAATCCTTGTTGATTTTGGGAACACTTAT -1291
人 HDGF2 核酸	- -----GGGAA-AAGACATTTTAGC-----CTTT-- -919
小鼠 HDGF 核酸	- CCCCTGACCCCAGGGTTCAAGGAATTGTAGTTTAACATCTAGACTTTGG -1341
人 HDGF2 核酸	- --TTTAAAAGTT----- -929
小鼠 HDGF 核酸	- AGTTTCCAAGTTTGGGCCTAGGACCTGGAGGGAGCTAAGAGCTGAAGAAT -1391
人 HDGF2 核酸	- --ACTGATTTAATTTCA-----TGT-TATTTGGTT-----GCATGAA-- -963
小鼠 HDGF 核酸	- CAACTGATTTGCATTGAGGAAATGTCTCTTTAGATCTCAGGGCAGAAATG -1441
人 HDGF2 核酸	- -----GTTGCCCTTAACCACT----AAGGATTAT---C -989
小鼠 HDGF 核酸	- ATAACCTGGGGAGACCTGCTGCCTTCATCTACTTCCAATGCTTGAGGCC -1491
人 HDGF2 核酸	- A-----AGATTTTGT-CGCAGACTTATA-----CATGTCT- -1018
小鼠 HDGF 核酸	- AGCCTGTAGTCAGATATTTACCCAGACATAAAGGAAAAGACCATTTTTT -1541
人 HDGF2 核酸	- --AGGATC -1024
小鼠 HDGF 核酸	- TTAGGAAATGTTTTTAATAAAA -1563

同一性: 68.7%

图 1(续)

人 HDGF2 蛋白	- MARP-RPREYKAGDLVFAKMKGYPHWPARIDELPEGAVKPPANKYPIFFF -49
小鼠 HDGF 蛋白	- MSRSNRQKEYKCGDLVFAKMKGYPHWPARIDEMPEAAVKSTANKYQVFFF -50
人 HDGF2 蛋白	- GTHETAFLGPKDLFPYKEYKDKFGKSNKRKGFNEGLWEIENNPVKFTGY -99
小鼠 HDGF 蛋白	- GTHETAFLGPKDLFPYEESEKFGKPNKRKGFSEGLWEIENNPTVKASGY -100
人 HDGF2 蛋白	- QAIQQQSSS-----ETEGEGGN-----TADASSEEEGDRVEEDGKGKRKN -139
小鼠 HDGF 蛋白	- QSSQKKSCAAEPEVEPEAHEGDGDKKGS AEGSSDEEG-KLVIDEPAKEKN -149
人 HDGF2 蛋白	- EKAGSKRKKSYSKSSKQSRKSPGDEDD----- -168
小鼠 HDGF 蛋白	- EKGTLKRRAGDVLEDSPKRPKESGDHEEEDKEIAALEGERHLPVEVEKNS -199
人 HDGF2 蛋白	- -----KDCKEEENKSSSEGGDAGNDTRNTTSDLQKTSEGT -203
小鼠 HDGF 蛋白	- TPSEPDSGQGPPAEEEEEGEEAAKEEAEAPGVRDH-----ESL -237

同一性: 53.7%
相似性: 9.4%

图 2